Observer weighting of monaural level information in a pair of tone pulses

Mark A. Stellmack and Neal F. Viemeister
Department of Psychology, University of Minnesota, Minneapolis, Minnesota 55455

(Received 17 September 1999; revised 4 January 2000; accepted 3 March 2000)

A correlational analysis was used to assess the relative weight given to the levels of two monaurally presented tone pulses for interpulse intervals (IPIs) ranging from 2–256 ms. In three different experimental conditions, listeners were instructed to discriminate the level of the first pulse, the level of the second pulse, or the difference between the levels of the two pulses. The level of the target pulse was chosen randomly and independently from trial to trial from a Gaussian distribution. The level of the nontarget pulse was either fixed at 75 dB SPL or varied in the same manner as the level of the target. In the tasks in which one pulse was to be ignored, listeners gave increasing weight to the nontarget component as IPI decreased. Listeners weighted the level information in the pulses appropriately only when the IPI approached 256 ms. When the listeners were instructed to compare the pulse levels to one another, two of three listeners weighted the levels optimally at all IPIs, while the third listener did so only at the longest IPI. For the two listeners who weighted the pulses optimally, a minimum in performance was achieved at IPIs around 16–32 ms. Intensity discrimination thresholds were also measured for one pulse in the presence of a second fixed pulse for IPIs of 2–256 ms. Thresholds were higher in all the two-pulse conditions relative to a one-pulse condition, and were dependent on the level of the nontarget pulse but not on IPI. The results indicate that level information is integrated to some extent over fairly long durations, but not in a manner that is consistent with simple temporal integration.
[S0001-4966(00)02506-6]

PACS numbers: 43.66.Mk, 43.66.Fe [SPB]

INTRODUCTION

Auditory information is integrated or accumulated over time, a process which is exhibited in a number of ways. In particular, detection thresholds decrease with increasing signal duration, and temporal resolution is limited by temporal integration, such that rapid changes that occur in an acoustical signal are combined or ''smoothed over'' by the auditory system. The ''resolution-integration paradox'' refers to the fact that studies of temporal resolution in the auditory system (e.g., Viemeister, 1979; Forrest and Green, 1987) generally produce estimates of around 3 ms for the integration time, while studies of the detection of signals as a function of duration yield integration times on the order of hundreds of ms (e.g., Green, 1960; Green and Swets, 1966). Often, this paradox is explained by assuming that there are two systems with different time constants, and that the system that is used depends on the particular listening task (Green, 1985; de Boer, 1985).

Viemeister and Wakefield (1991) proposed a multiple-looks hypothesis to attempt to account for long-term temporal integration in a system with a short integration time. The idea is that the auditory system accumulates many short-term ''looks'' at the signal, retains those looks in memory, and selectively uses them as the basis for decisions. In order to ''selectively use'' the looks, a listener might give equal consideration to the information in all looks (e.g., in detecting a long-duration signal), the listener might only attend to particular looks while ignoring or discarding others (e.g., in detecting a signal at a particular temporal position in a noisy background), or the listener might compare looks to one another (e.g., in identifying a temporal pattern of loudness changes). (These hypothetical listening strategies represent the ideal strategy for each situation.) The manner in which the looks are utilized is given by a weighting function that describes the relative influence of each look on a decision. In order to account for actual temporal integration data for the detection of long-duration stimuli, Viemeister and Wakefield proposed that listeners apply a weighting function in which early portions of the signal are given relatively little weight, while the relative weight increases for later portions of the signal. Consistent with the notion that a long-duration stimulus is effectively processed as many short-term looks, Viemeister and Wakefield found that detection thresholds for a signal consisting of two 10-ms tone pulses that are 100 ms apart are unaffected by the level of a 50-ms noise burst temporally centered between the tone bursts.

Buus (1999) examined the multiple-looks hypothesis using an analysis in which he computed the perceptual weight given to a number of discrete short-duration pulses that comprised a signal to be detected. In a two-interval forced-choice task, the signal consisted of 25-ms, 1-kHz tone pulses (Buus *et al.*, 1996). On each trial, the levels of the signal pulses in the signal interval were selected randomly and independently from a distribution with a particular mean level and a standard deviation of either 3 or 6 dB, in separate conditions. In order to compute the weight given to a particular pulse, or temporal look, the data were sorted into ten categories based upon the level of that look on each trial. Then d'^2 was plotted as a function of the average squared intensity of the looks in each of the ten categories. This was done separately for each temporal look. Thus for a signal consisting of six temporal segments, six different plots of d'^2 vs intensity-squared were produced. Buus showed that the squared perceptual weight given to a particular look is proportional to the slope of the line through these data for the given look. In other words, if a particular look is given a large perceptual weight, its level on a given trial will have a large influence on the detectability of the signal as a whole, and d'^2 will vary to a large extent with intensity-squared, resulting in a relatively large slope value for the line relating d'^2 to intensity-squared. On the other hand, if a particular look is given no perceptual weight, then detectability of the signal will not depend upon the level of that look on any given trial, and d'^2 will be unrelated to intensity-squared (i.e., the function will have a slope of zero).

Buus found that detection of the signal was based on an optimal combination of the information in the individual temporal looks. Each look was given approximately equal weight by the listeners in detecting the signal, which is appropriate given that each look was equally informative overall because each look had the same distribution of levels across trials. This result suggested that detectability of a long-duration signal is based upon an optimal combination of the information in a number of shorter segments of the signal, a result that Buus concluded was consistent with the general form of the multiple-looks hypothesis described by Viemeister and Wakefield (1991). It should be noted, however, that equal weights also would be computed for the individual pulses of the signal if energy were integrated over the entire duration of the signal, that is, if the entire duration of the signal actually were treated as a single ''look'' by the auditory system. (In addition, somewhat different results were obtained in a separate condition in which a second off-frequency masker was comodulated with the on-frequency masker.) Similarly, in a sample-discrimination paradigm, Lutfi (1990) asked listeners to discriminate changes in the frequency, level, or duration of a sequence of pulses. Using a COSS (conditional on single stimulus) analysis to compute observer weights (Berg, 1989), Lutfi also measured essentially equal weights for all pulses, which for this task, as in the detection task of Buus (1999), is the optimal weighting strategy.

One implication of the multiple-looks hypothesis is that the listener accumulates information about a number of short-term looks in memory and the listener then can access selectively the information in the individual looks depending on the task at hand. In other words, while Lutfi (1990) and Buus (1999) showed that the listener gives equal weight to the individual looks in a situation where all looks contribute equally to the ''signal,'' the multiple-looks hypothesis would also predict that the listener can ignore the information in uninformative temporal segments of a stimulus. For example, as described earlier, Viemeister and Wakefield (1991) showed that detection thresholds for a signal consisting of two 1-kHz tone pulses were unaffected by increases or decreases in the level of noise during the time between the two pulses. This seems to support the notion that relatively short, temporally discrete elements of a stimulus can be used or ignored by the listener as necessary.

One question that arises is: In a task in which not all temporal segments of the stimulus are relevant, does the pattern of perceptual weights indicate that listeners are able to ignore the irrelevant information? Gilkey and Robinson (1986) computed weights for individual temporal segments of a stimulus interval in which listeners were instructed to detect a 500-Hz tonal signal in a broadband noise masker. The 100-ms signal was approximately temporally centered in the 148-ms masker, so that in a signal-bearing interval (of the single-interval, yes–no task) there was a masker fringe preceding and trailing the signal. This allowed the authors to compute weights for temporal segments of the stimulus that did not contain the signal as well as those that contained both signal and noise. Generally, it was observed that listeners either gave little weight to the leading masker fringe, indicating that they ignored the initial temporal segment of the stimulus, or listeners gave slightly negative weight to the leading masker fringe, indicating that they compared the noise-alone segment to later segments that might potentially contain the signal. For all listeners, the perceptual weights rose during the first half of the signal, while the weights were inconsistent across listeners for the second half of the signal. Rather large weights were computed for some listeners for the final segment of the interval, which also always consisted of masker alone (the trailing fringe). These results give some indication that listeners use, or are not able to ignore, temporal portions of the stimulus that never contain the signal. In addition, the weights computed by Gilkey and Robinson gradually changed during the duration of the stimulus, while Buus (1999) computed weights that were essentially equal for the entire stimulus duration. The source of these differences is not clear. The maskers in both experiments were of longer duration than the signals, resulting in temporal fringes preceding and following the signals. One difference between these studies is that Gilkey and Robinson used a 100-ms tone as the signal while Buus used a number of 25-ms tone pulses which essentially resulted in an amplitude-modulated signal. Furthermore, the weights computed by Buus were based on the intensity of only the signal in each temporal segment of the stimulus. The intensity of the masker in each temporal segment, an additional source of variability in the computation of weights, was ignored.

Lutfi (1992) performed a sample-discrimination experiment in which only one tone of a 10-tone sequence was identified as the target. A different sequence was presented in the two intervals of each trial, and listeners were instructed to select the interval in which the level of the target was chosen from an incremented distribution of levels. In a separate condition, listeners made similar judgments about the frequencies of the tones. Each tone in the sequence was 30 ms in duration, rectangularly gated, with no gaps between the tones. In this task, the optimal weighting strategy is one in which the listener gives no weight to the nine nontarget tones. Lutfi found that the specific form of the weighting function depended upon the relative jitter applied to the target and nontarget tones. The target tone was given the greatest weight when its jitter was larger than that of the nontarget tones, and all tones were weighted nearly equally when the jitter of the target tone was smaller than that of the nontarget tones. In most cases, many of the nontarget tones were given nonzero weight. There was a tendency for the final tone in the sequence, a nontarget tone, to be given slightly greater weight than the other nontarget tones.

Another type of task for which the optimal weighting strategy would involve unequal weights over time is one in which the listener is required to make judgments as to the relative intensity of temporal segments of a signal, or, equivalently, a task requiring the listener to judge the temporal order of pulses with different levels (Hirsh, 1959; Babkoff and Sutton, 1963; Ronken, 1970; Green, 1973). In one such experiment, Green (1973) asked listeners to judge the temporal order of two 1-ms tonal pulses that differed in intensity by 10 dB. He measured percent correct as a function of the temporal separation of the pulses for frequencies of 1, 2, and 4 kHz. Green found that for all frequencies performance peaked at separations around 2–4 ms, decreased as the separation increased from 4 to 32 ms, then increased once again as the separation was made larger than 32 ms. Similar results were reported by Babkoff and Sutton (1963). Green concluded that the nonmonotonic change in performance with increasing temporal separation indicates that two distinct temporal processes are involved in the discrimination of the intensity of sequentially presented tone pulses.

Schlauch *et al.* (1999) performed a series of experiments designed to assess the ability of listeners to access level information in discrete temporal elements of an auditory stimulus. In each trial, a train of either three or four noise bursts was presented to listeners. The levels of successive bursts alternated between 60 and 80 dB SPL, with the level of the first burst chosen at random. Listeners were instructed to indicate whether the first burst was a low- or high-level burst, and performance was measured as a function of the presentation rate of the bursts. In a second experiment, listeners were instructed to count the number of bursts in a train of equal-level noise bursts that varied in their rate of presentation from trial to trial. No feedback was given in either experiment. Schlauch *et al.* reported that both experiments yielded similar results: listeners were able to perform both tasks for low stimulus presentation rates (large temporal spacings), but performance dropped as presentation rate increased (temporal spacing decreased). The authors estimated the upper limit of temporal resolution as measured by these tasks as about 20 Hz, corresponding to a temporal separation (silent period) of about 25 ms, which is in good agreement with the data on temporal order judgment discussed above.

Experiment 1 examines in more detail the issue of the perceptual weight given to discrete temporal elements of an auditory stimulus, more specifically, the weighting strategies adopted by listeners when a stimulus is composed of a number of temporally discrete pulses and listeners are asked to attend to the pulses in a number of different ways. The question to be addressed is: Do listeners alter their weighting strategies in an intensity discrimination task when the relevance of short-duration segments of the stimulus is varied?

I. EXPERIMENT 1: PERCEPTUAL WEIGHTS

In one set of conditions of this experiment, listeners are asked to discriminate changes in the level of one temporal segment of the stimulus (the target) while ignoring level changes in another temporal segment (the nontarget). In a case such as this, the optimal weighting strategy is one in which all perceptual weight is given to the target and a weight of zero is given to the nontarget. In another condition, listeners are asked to compare the level of the first pulse to the second pulse and to indicate which pulse has a higher level. In this case, the optimal weighting strategy is one in which weights of equal magnitude but opposite sign are given to the pulses. Weighting strategies are assessed as a function of the temporal separation between the pulses. Feedback is provided to listeners in each condition.

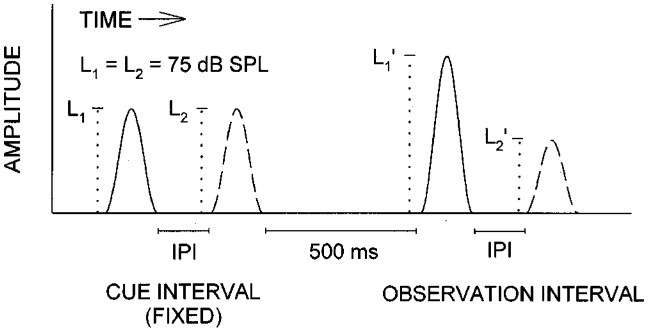

FIG. 1. A representation of a trial in which the levels of both the first and second pulses were varied. The first interval (the cue) consisted of two pulses with levels of 75 dB SPL. The second interval (the observation interval) consisted of two pulses with randomly and independently chosen levels. (All *L*s are in units of dB SPL.) The IPI was fixed for an entire block of trials. Because the levels of the pulses were selected randomly and independently from trial to trial in certain experimental conditions, the changes in level of the two pulses (relative to the cue) might be consistent with or opposite to one another on any given trial in those conditions.

A. Methods

All stimuli were composed of tone pulses (described below) and were presented over headphones to the listener's left ear. Each trial, an example of which is shown in Fig. 1, consisted of two intervals (resulting in a cued single-interval task). In the first interval (the cue), the first and second pulses were both presented at their mean level. The pulses in the cue were separated by the interpulse interval (IPI) which was determined by the specific condition that was being run. The second interval (observation interval) consisted of two pulses that were separated by the same IPI as the cue. The IPI for the cue and observation interval was the time between the offset of the first pulse and the onset of the second pulse. Data were collected for IPIs of 2, 8, 16, 32, 64, 128, and 256 ms.

The first and second pulses in the observation interval had levels that were randomly and independently chosen or that were fixed at the mean level on each trial, depending on the particular experimental condition. When the levels were randomly chosen, they were chosen from a Gaussian distribution with a mean of 75 dB SPL. Thus when the levels of both the first and second pulses were randomly chosen, the pulses might change level in the same or opposite directions relative to the cue on a trial-by-trial basis.

All pulses were 1-kHz pure tones, 10 ms in duration, with 5-ms raised cosine on–off ramps (0-ms steady state). The level of each pulse was defined to be the level in dB SPL of a continuous tone with the same peak amplitude. The cue and observation interval of each trial were separated by 500 ms of silence. The IPI and instructions to the listener were fixed during each 50-trial block. Twenty blocks were presented to listeners for every IPI in each experimental condition so that 1000 responses were obtained for each. In all of the conditions described below, feedback that was consistent with the particular experimental condition was given to the listener after each trial.

Listeners were instructed to use the information in the first and second pulses in a particular way in each experimental condition. In one set of conditions, the first pulse was identified as the target. In this case, the listener was instructed to indicate whether the first pulse increased or decreased in level relative to that of the cue by pressing one of two keys on a computer keyboard. The listener was instructed to ignore the second pulse in both intervals when the first pulse was the target. In one condition, the level of the second pulse in the observation interval was fixed at its mean level (75 dB SPL), and in a separate condition the level of the second pulse was selected randomly from a Gaussian distribution with the same parameters as that of the first pulse. In this latter condition, the first and second levels were selected independently. For convenience, these separate conditions were called ‘‘target-1/fixed-2’’ and ‘‘target-1/variable-2.’’

In another set of conditions, the second pulse was identified as the target and the listener was instructed to indicate whether it increased or decreased in level relative to the cue while ignoring the first pulse. As above, the nontarget (first pulse) level was fixed at its mean level in one condition (‘‘target-2/fixed-1’’), and in a separate condition (‘‘target-2/variable-1’’) the nontarget level in the observation interval was selected from a Gaussian distribution with the same parameters as that of the target.

In a final condition, the difference condition, the listeners were instructed to compare the two pulses in the observation interval to one another and to indicate whether the second pulse was higher or lower in level (‘‘louder’’ or ‘‘quieter’’) than the first pulse. In this condition, the levels of the two pulses in the observation interval were always selected randomly and independently. The fixed-level cue, consisting of two pulses at the same IPI as the observation interval, was presented in this condition even though it may not have been particularly useful to the listener. This was done so that the stimuli in this condition were statistically identical to those in the target-1/variable-2 and target-2/variable-1 conditions. The difference condition was run at the same IPIs as the previous conditions.

In all conditions, when a pulse’s level varied from trial to trial, the level was selected randomly from a Gaussian distribution with a mean of 75 dB SPL and a standard deviation (in dB) chosen to yield approximately 90% correct for the baseline intensity discrimination task for a single pulse in isolation, determined by trial and error. To achieve this level of performance, the standard deviation was set to 4 dB for listener S1 while it was set to 8 dB for the remaining four listeners. For conditions in which the levels of both the first and second pulse were variable, the levels were sampled independently from the Gaussian distribution. The levels were limited to a range of ± 2.5 standard deviations from the mean in order to avoid excessively high levels, producing overall ranges in level of 20 dB for listener S1 and 40 dB for the remaining four listeners. If the randomly chosen levels would produce a stimulus for which a correct response was undefined (0-dB change in level for the target, or equal levels for the two pulses in the difference condition), new levels were randomly chosen for the pulses.

As a baseline measure of performance, percent correct was also measured for a single pulse. In this case, the cue (a single pulse at the mean level) was presented in the first interval of each trial, followed by a single pulse with its level selected in the same manner as those in the variable level conditions described above. Listeners were instructed to indicate whether the second pulse was higher or lower in level than the cue.

All stimuli were generated digitally on a NeXT computer and converted to analog signals using the NeXT’s 16-bit D/A converters set to a rate of 44.1 kHz. The pulses were presented monaurally over Sony MDR-V6 headphones to listeners seated in an IAC sound-attenuating chamber. Listeners entered their responses on the computer keyboard and visual feedback was provided after each trial via the computer monitor.

Listeners were run during 2-h sessions during which approximately 15–20 50-trial blocks were completed. As stated previously, 1000 trials were run for each IPI in each condition. All trials of a particular condition were run before trials of another type were presented.

There were five listeners in the target-1 and target-2 conditions, and three of those listeners performed in the difference condition. One listener was the first author (S1). All remaining listeners were undergraduate volunteers who were paid to participate in the experiment. The first author was experienced in listening to these types of stimuli and required little training in order to become familiar with the task. The remaining four listeners had little or no experience in these types of listening tasks and were run for several thousand trials before data collection began. In addition, listeners were run in one or more blocks of practice trials in each condition before data were collected. All listeners ran the conditions in the following order: target-1/fixed-2, target-1/variable-2, target-2/fixed-1, target-2/variable-1. Within each condition, stimuli were presented from the longest IPI to the shortest IPI. The difference condition was run last, also from the longest to shortest IPI. Several randomly chosen conditions were repeated at the end of the experiment and no changes in performance or weighting strategies were observed, indicating that there were no order effects present.

B. Results

1. Computation of observer weights

The data in the present experiments were analyzed using a correlational analysis (Richards and Zhu, 1994; Lutfi, 1995). In this model, the listener’s responses are based on a decision variable that is the weighted sum of the changes in level of the individual components (in this case, the first and second pulses),

$$D = w_1(\Delta L_1) + w_2(\Delta L_2) + \varepsilon, \tag{1}$$

where the w's are the observer's weights for the level changes between intervals (in dB) of the first and second pulses (as indicated by the subscripts), and ε is an additive error term encompassing all variability unaccounted for by the weighted levels. The value ΔL for each pulse is the change in level of each pulse across intervals (e.g., $\Delta L_1 = L_1' - L_1$ from Fig. 1), or, equivalently, the difference between each pulse and the mean pulse level. Responses are assumed to be based on the value of D such that listeners respond

$$\begin{aligned} &\text{``quieter'' if } D<k; \text{ ``louder'' if } D>k, \\ &\text{``quieter'' or ``louder'' randomly if } D=k, \end{aligned} \tag{2}$$

where k is the value of the criterion adopted by the listener. In the absence of response bias, k will be equal to zero, the mean change in level of each pulse across all trials. Equation (2) represents rescaling of the continuous internal decision variable D as a dichotomous response variable. The relative weight given to the level information of a given component [in Eq. (1)] is proportional to the correlation between the trial-by-trial levels for that component and the listener's binary responses.

In the results reported here, the *magnitudes* (absolute values) of the correlation coefficients (rather than the actual correlation coefficients) were normalized to sum to 1.00, and the *sign* of each original correlation coefficient was retained. This was done in order to deal with situations in which negative correlations were obtained. When negative correlations are obtained, normalizing the actual correlation coefficients produces weights beyond the range of −1.00 to +1.00. Normalizing the magnitudes of the correlation coefficients more accurately reflects the relative weight given to each component across conditions. The sign of each weight then simply indicates whether the listener's responses were consistent with (+) or opposite to (−) the direction of the level change of that component.

2. Interpretation of observer weights

One can consider the weighting strategy used by an ideal observer that performs the task optimally, that is, a ''noiseless'' observer that maximizes percent correct. In these experimental tasks, every trial has an unambiguously correct response based on either the level change of the target or the difference between the levels of the two pulses. Thus an ideal noiseless observer that weights the information optimally for a given task will achieve 100% correct. When a human observer fails to achieve 100% correct, one can evaluate the extent to which imperfect performance results from (1) inappropriate weighting of the levels of the two pulses or (2) the influence of additional sources of variability other than the sum of the weighted levels.

The listener is said to use a nonoptimal weighting strategy when the relative weights for that listener differ from the ideal weighting strategy for a given condition. Given the method of normalizing the observer weights used here, the weights will fall in the range from −1.00 to +1.00. When the first or second pulse is identified as the target and the listener is instructed to ignore the nontarget pulse, the ideal weighting strategy (which maximizes percent correct) is one in which the observer gives the target pulse a weight of +1.00 and the nontarget pulse a weight of 0.00. Such a weighting strategy will produce 100% correct for the ideal noiseless observer. To the extent that weight is given to the nontarget pulse, performance will fall below 100% correct. When the listener is instructed to respond to the level of the second pulse relative to that of the first (the difference condition), the ideal weighting strategy is one in which the pulses are weighted equally but with opposite signs, that is, the weight for the second pulse is +0.50 and the weight for the first pulse is −0.50. This weighting strategy will produce 100% correct for an ideal observer in the difference condition.

Even if the human observer uses the optimal weighting strategy, performance (percent correct) will be adversely affected by the presence of other sources of variability [ε in Eq. (1)] that are unrelated to the sum of the weighted levels, such as the discriminability of the level changes in the individual pulses or interference in the memory trace of one pulse produced by another pulse. However, the error term cannot be computed directly because the listener's responses are the result of generating dichotomous responses using Eq. (2) from values of the continuous variable D in Eq. (1).

Although the error term cannot be computed directly, its relative influence across experimental conditions can be estimated indirectly. An estimate of the relative influence of ε in Eq. (1) indicates how well the weighted combination of levels alone accounts for the listener's responses. One way in which the relative influence of the error term across conditions can be estimated is by computing the percentage of the listener's responses that can be predicted by the listener's weights.[1] In order to calculate this percentage, the computed weights were inserted into Eq. (1) and ε was set to zero. Simulated responses were obtained by inserting into Eq. (1) the actual levels that were presented to the listener on each trial and using Eq. (2) to generate a response.[2] The percentage of simulated responses that agreed with the listener's actual responses (the percentage of responses predicted by the weights) was then computed. This value is inversely related to the relative contribution of ε in the decision process. If $\varepsilon = 0$, then 100% of the responses can be predicted by the weights [when Eq. (2) is used to generate dichotomous responses from the continuous variable D]. As ε becomes large, the percentage of responses predicted by the weights will drop to a lower limit of 50%. Furthermore, the particular weighting strategy that is adopted by the listener becomes irrelevant when the weights predict a very low percentage (near 50%) of the responses. In such cases, the weights have essentially no predictive value (responses are uncorrelated with the levels of either pulse). A comparison of the influence of ε across conditions indicates which conditions are influenced to a greater or lesser extent by factors other than the weighted level changes. In practice, it is unlikely that the

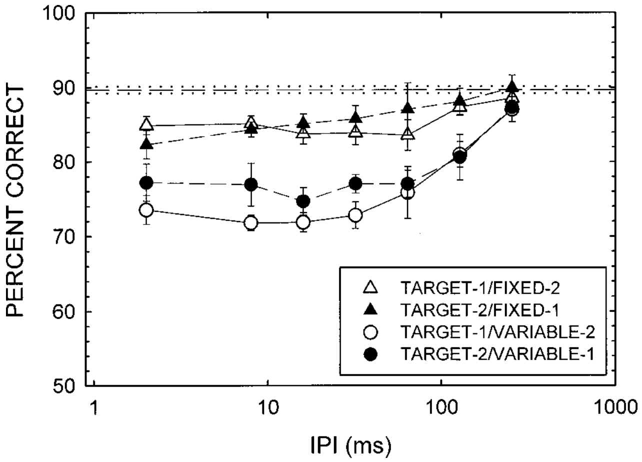

FIG. 2. Percent correct as a function of IPI in the two target-1 conditions and the two target-2 conditions. Percentages have been averaged across five listeners. Error bars represent standard errors of the mean. The dashed line at the top of the figure represents mean percent correct for the same five listeners in the single-pulse condition. The dotted lines above and below the dashed line represent ±1 standard error of the mean.

weights will account for 100% of the responses in a given condition simply because the levels of the pulses are chosen randomly from trial to trial and there is some probability in each trial that the level of a particular pulse will be indiscriminable from its mean level. (See Stellmack *et al.*, 1999 for a more detailed discussion of the interpretation of the weights.)

3. Target-1 and target-2 conditions

Because the patterns of results for the five listeners were quite similar in the target-1 and target-2 conditions, the results were averaged across listeners. Figure 2 shows percent correct as a function of IPI in all of the target-1 conditions (open symbols) and target-2 conditions (filled symbols). Triangles represent the conditions in which the level of the nontarget pulse was fixed, and circles represent the conditions in which the level of the nontarget pulse varied. The error bars indicate standard errors of the mean. The dashed and dotted horizontal lines near the top of the panel indicate the average percent correct and ±1 standard error of the mean, respectively, for these listeners in the baseline (single pulse) condition.

On average, percent correct was lower when two pulses were present relative to performance in the single-pulse condition except for the longest IPI (256 ms), where performance was essentially equal across conditions. In addition, there was a substantial decrement in performance when the level of the nontarget pulse varied from trial to trial (circles) relative to conditions in which the level of the nontarget pulse was fixed (triangles) for all but the longest IPI.

In the top panel of Fig. 3, percent correct scores from the target-1/variable-2 condition (open symbols) and target-2/variable-1 condition (filled symbols) have been replotted from Fig. 2. The middle panel of Fig. 3 shows the average normalized target weights as a function of IPI for the same conditions. (For the conditions in which the nontarget level is *fixed*, the relative weights are undefined.) Because the weights were normalized to sum to 1.00, the nontarget weights are simply 1.00 minus the target weights that are plotted in the middle panel of Fig. 3. Although negative weights are possible, all average weights for both the target and nontarget pulses were positive in these conditions. The bottom panel of Fig. 3 shows the percentage of responses predicted by the weights for the same conditions. The ordinate in the bottom panel has been inverted to reflect the inverse relationship between the percentage of responses predicted by the weights and the relative influence of variance due to factors other than the weighted sum of the pulse levels. Thus higher relative amounts of unexplained variance are represented toward the upper portion of the panel.

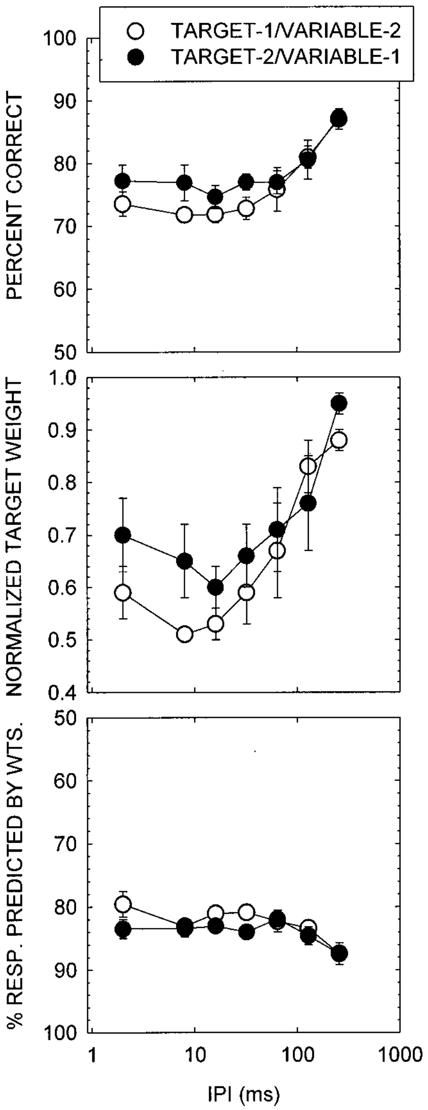

FIG. 3. In the top panel, percent correct is replotted from Fig. 2 for the target-1/variable-2 and target-2/variable-1 conditions. For the same conditions, the middle panel shows the target weight and the bottom panel shows the percentage of responses predicted by the weights, both as a function of IPI. All data are averaged across five listeners. Error bars represent standard errors of the mean.

The pattern of weights in the middle panel of Fig. 3 is similar to the pattern of performance shown in the top panel. High percent correct values in the top panel correspond to high target weights in the middle panel, with a similar correspondence for lower values. Percent correct was slightly higher for the target-2 condition (filled symbols) relative to the target-1 condition (open symbols) for IPIs of 32 ms and less, and the same is true for the average target weights. In contrast, the percentage of responses predicted by the weights (in the bottom panel) remains relatively constant across IPI with a slight increase at the largest IPI. In other words, *relative* performance across these conditions is more closely associated with the relative weights given to the two pulses and is not influenced to a great extent by differences in the amount of additional unexplained variance across IPI.

Recall that the optimal weighting strategy that would produce the maximum percent correct is one in which the target is given a weight of 1.00. The target weight approaches but does not reach 1.00 at the largest IPI, indicating that changes in the level of the nontarget pulse influenced responses at all IPIs. As IPI increases from about 16 ms, the target weight becomes larger (closer to optimal). Target weights were at a minimum for IPIs around 8–16 ms. For these intermediate IPIs, the greatest nontarget weights were computed. The weight given to the target was slightly higher at the shortest IPI (2 ms).

4. Difference condition

In the difference condition, two pulses with randomly and independently chosen levels were presented in the observation interval of each trial, and listeners were instructed to indicate whether the level of the second pulse was higher or lower than that of the first pulse. The results for the difference condition were less consistent across listeners than those for the previous conditions, so the data are presented individually in Fig. 4. Data for listeners S1 and S2 were qualitatively similar to one another, and are plotted in the left-hand column of Fig. 4 (in the same format as the data shown in Fig. 3). Data for listener S3 were somewhat different and are plotted separately in the right-hand column. In the middle panel of each column, the filled symbols are weights for the second pulse and the open symbols are weights for the first pulse. Different symbol types represent different listeners, as shown in the figure legend. Due to unavailability, S3 did not provide data for the smallest IPI (2 ms).

For S1 and S2 (left-hand column of Fig. 4), percent correct followed the same pattern in the difference condition as in the previous conditions (shown in Fig. 3). Best performance was obtained at the largest IPI (256 ms) and a minimum was observed at intermediate IPIs (32 ms for S1, 16 ms for S2). Recall that the optimal weighting strategy for the difference condition is one in which the weights are equal in magnitude but with opposite sign (+0.50 and −0.50). In contrast to the results for the target-1 and target-2 conditions, the weighting strategies for these two listeners (middle panel) remained fairly constant on average and nearly optimal across IPIs (although the second pulse was given slightly greater weight, as shown by the relative magnitudes of the weights), while the variation in the percentage of responses predicted by the weights (bottom panel) was more similar to the variation in percent correct. As the relative amount of unexplained variability increased, percent correct decreased. It appears that in the difference condition, performance across IPI was less dependent upon the particular weighting strategy adopted by the listener and more influenced by additional, unexplained sources of variability than in the target-1 and target-2 conditions.

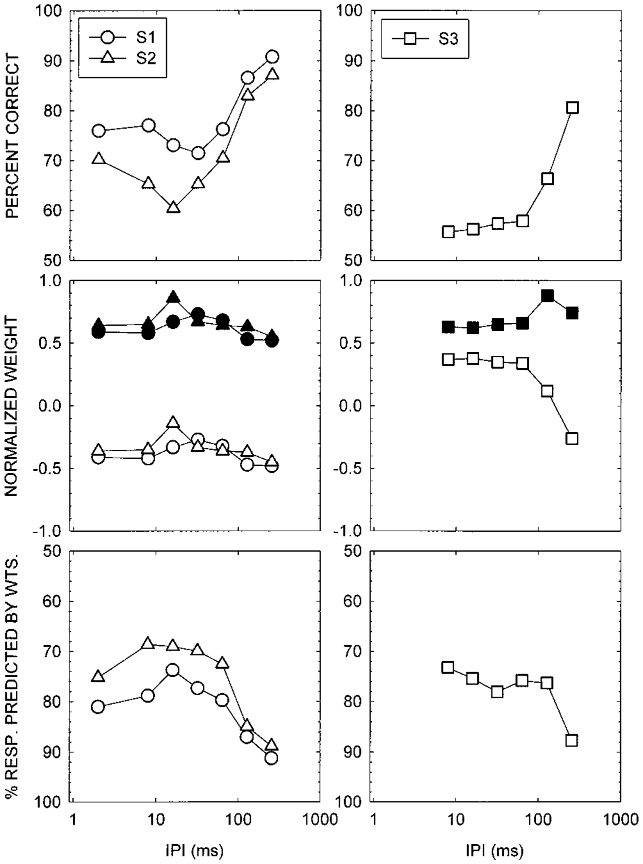

FIG. 4. The results of the difference condition, plotted in the same format as Fig. 3. Results are shown for listeners S1 and S2 in the left-hand column and for listeners S3 in the right-hand column. In the middle panels of both columns, open and filled symbols represent weights for pulse 1 and pulse 2, respectively. Each symbol type represents results for a different listener, as indicated in the legend.

Listener S3 (right-hand column of Fig. 4) provided a slightly different pattern of results. Percent correct was at a maximum for the largest IPI, as for the other two listeners, and quickly dropped to near chance as IPI decreased. Weights of opposite sign were obtained only for the largest IPI, although slightly larger weight was given to the second pulse in all conditions, as was seen for the other two listeners. It appears that at the shorter IPIs, S3 had substantial difficulty with the difference task and adopted a strategy of merely responding to the change in overall level between intervals, with the second pulse weighted slightly more heavily.

C. Discussion

The results indicate that, over the range of IPIs tested here, listeners were unable to completely ignore an adjacent temporal segment of a stimulus that was qualitatively similar to the target segment when performing an intensity discrimination task. For 10-ms, 1-kHz pulses separated by intervals from 2 to 256 ms, the nontarget pulse influenced the responses of the listener. Listeners appeared to have the most difficulty ignoring the nontarget pulse at intermediate IPIs (around 8–16 ms), with better performance at longer and shorter IPIs. The target weight was slightly greater when the second pulse was the target (compared to when the first pulse was the target).

The difference task is similar to the tasks requiring temporal order judgments that were described in the Introduction, although here the levels of the pulses were randomized from trial to trial. For listeners S1 and S2, there is some indication that percent correct is lowest for IPIs around 16–32 ms, consistent with the results of Babkoff and Sutton (1963) and Green (1973), but the nonmonotonicity in performance across IPIs is not nearly as pronounced as in the previous studies. Green (1973) observed that the nonmonotonicity was most evident in the data of unpracticed subjects. The large number of trials run in the present study may have served to reduce this effect.

Note that the stimuli in the difference condition were statistically identical to the stimuli in the target-1 and target-2 conditions with variable nontargets. Listeners S1 and S2 (and S3 to some extent) altered their weighting strategies depending upon the specific task at hand. Slightly greater positive weight was given to the target pulse in the target-1 and target-2 conditions, with relatively small positive weight given to the nontarget (Fig. 3). In the difference condition, weights of more similar magnitude but opposite sign were given to the two pulses (Fig. 4). Although the weights for only two of the three listeners varied in a manner consistent with the task instructions across all IPIs, this provides support to the notion that the levels of the individual pulses were treated as individual ‘‘looks’’ by the auditory system. If the level information were completely integrated across pulses, the weights would have been approximately equal and with the same sign across all conditions.

In the target/variable conditions, relative performance across IPIs was primarily determined by the weighting strategy adopted by the listeners. Decreases in percent correct were associated with increases in the weight given to the irrelevant nontarget pulse. In contrast, for two of the three listeners in the difference condition, relative performance was not strongly related to the listeners' weighting strategy. The weighting strategy was nearly fixed and close to optimal across IPI. Decreases in performance were mirrored by increases in additional, unexplained variability in responses. (For the third listener in the difference condition, performance declined as both the weighting strategy became nonoptimal and the variability increased.) Thus it appears that listeners adopt a weighting strategy that is closer to optimal when they are required to make decisions based on the entire stimulus rather than on certain temporal segments of the stimulus.

Green (1973) suggested that performance in the discrimination of temporal order exhibits a nonmonotonicity because two different mechanisms are at work at short and long IPIs, with neither mechanism extremely effective at intermediate IPIs. At very short IPIs, listeners perceive a slight qualitative difference between stimuli consisting of a soft–loud tone sequence versus a loud–soft sequence. Apparently, with brief IPIs listeners perceive the two pulses as a single auditory event, with the change in loudness across the pulses producing a qualitative difference depending on the direction of the change. (Note that there are no long-term spectral differences between the time-reversed soft–loud and loud–soft tone sequences used by Green when one considers the spectrum of the entire tone sequence.) At intermediate IPIs, these qualitative differences become more difficult to hear as the pulses become more distinct. At long IPIs, the listener can clearly perceive two distinct pulses and the ordering of their intensities can be easily followed.

The idea that intensity information in a sequence of tone pulses is processed as a number of short-term looks does not imply that the listener necessarily perceives the pulses as individual auditory events. The fact that weights of opposite sign and nearly equal magnitude were obtained for two listeners at short IPI's indicates that the levels of the two pulses effectively were processed independently. The effect is similar to that in situations involving the detection of amplitude modulation at high modulation frequencies. Although ‘‘roughness’’ might be perceived rather than discrete fluctuations in intensity (Wright and Dai, 1999), the fact that the fluctuations can be detected at all indicates that at some level of processing the fluctuations are resolved by the auditory system. When the task requires listeners to selectively process a particular peak in the fluctuating waveform, as in the present target/variable task or Schlauch *et al.* (1999), the inability to perceptually segregate the fluctuations limits their ability to do so.

One reason that listeners may have tended to perceptually fuse the stimuli in the present experiment at short IPIs is that the temporal segments of the signal were qualitatively similar (having the same frequency and duration). It is possible that listeners would be more able to selectively process the temporal segments of a stimulus if they were more qualitatively dissimilar, for example, if they were tones of different frequencies or durations. The implication is that interactions in the processing of intensity information in the individual pulses occur at some more central level of processing, after the intensities of the individual looks are extracted at more peripheral levels. This is, in effect, the basic premise of the multiple-looks hypothesis.

The results also may have been influenced by the type of cue that was used, specifically, two pulses fixed in level at the mean level of the variable pulses. Perhaps the two-pulse cue influenced the listeners to perceptually combine the two pulses in both intervals even at the larger IPIs. Perhaps a one-pulse cue would encourage the listener to hear out the target pulse more readily, or perhaps presenting only the observation interval on each trial with no cue would produce such results.

To examine these possibilities, some of the conditions of experiment 1 were repeated with no cue and with a cue consisting of only one pulse at the mean pulse level. The target-1/variable-2 conditions of experiment 1 were repeated with these different cues for IPI = 2, 32, and 256 ms. The results are shown in Fig. 5 for listeners S1 and S4 in the same format as Fig. 3. The observation interval of each trial was identical to those of experiment 1, while the cue consisted of either one pulse (filled squares) or two pulses (‘‘time 2,’’

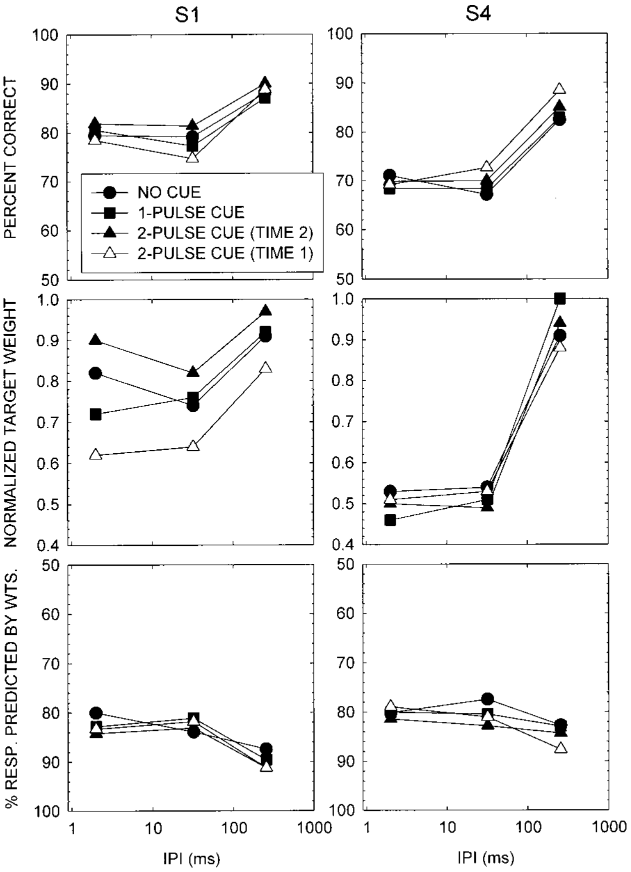

FIG. 5. Data from two listeners for selected conditions of experiment 1, repeated with different types of cues, as indicated in the legend. Results are plotted in the same format as Fig. 3. Data for the two-pulse cue condition at time 1 were gathered for experiment 1; all remaining data were collected at a later date.

filled triangles), or the observation interval was presented with no cue (filled circles). The two-pulse-cue condition was rerun because these data were gathered several months after the original data. The original two-pulse-cue data for these listeners from experiment 1 (''time 1,'' open triangles) are replotted for comparison.

It appears that the specific type of cue, including the complete lack of a cue, has little effect on the results. For S4, the results for no cue and a one-pulse cue are essentially the same as those for the two-pulse cue at both times 1 and 2. S1 produced the highest target weights for the two-pulse-cue condition at time 2, and identical, slightly lower target weights for the no-cue and one-pulse-cue conditions at IPI $=32$ and 256 ms. All of the target weights for S1 measured at time 2 were slightly larger than those measured at time 1, representing a practice effect similar to that documented by Leek and Watson (1984). This type of practice effect was not seen for S1 when some of the conditions were repeated at the time of the initial data collection. In spite of the larger target weights over time for S1, both listeners still show the same general result as the original data: the largest target weights were obtained for IPI$=256$ ms with substantially lower target weights at IPI$=2$ and 32 ms.

A result that is not predicted by the multiple-looks hypothesis is the observation that a second nontarget pulse affected intensity discrimination performance even when the level of the nontarget was fixed (Fig. 2, triangles). In this condition, listeners were required simply to indicate the direction of a level change occurring during the observation interval without the possible additional burden of ignoring a potentially conflicting level change, or of attributing two level changes to their respective temporal positions. The decrease in overall performance in the target/fixed conditions suggests that just-noticeable differences (JNDs) for intensity may be poorer (larger) for a tonal pulse in the presence of a second, fixed tonal pulse of the same frequency and duration. Such an effect was observed by Plack *et al.* (1995). An increase in intensity JNDs in the presence of a fixed nontarget would be predicted if intensity information were simply integrated across the two pulses, but the patterns of weights seen in the various conditions described above, in particular the fact that different patterns were obtained depending on the specific task, suggests that level information in the two pulses was processed independently, at least to some extent. To explore this issue further, an additional experiment was performed to examine the effect on intensity JNDs of a fixed nontarget pulse.

II. EXPERIMENT 2: INTENSITY JNDs

In this experiment, intensity JNDs are measured for a tonal pulse in the presence of a second, temporally discrete tonal pulse with a fixed level. Both pulses have the same frequency and duration. Intensity JNDs are measured as a function of the time between the two pulses (interpulse interval, IPI), ranging from 2 to 256 ms. JNDs are measured in conditions in which the level of the fixed pulse is -10, 0, or 10 dB relative to the pedestal level of the target pulse (in separate conditions). Intensity JNDs are also measured for a single pulse in isolation. If intensity simply is integrated across the pulses, then, relative to the single-pulse condition, intensity JNDs must increase by 0.41, 3.0, and 10.4 dB $[10\log(\Delta I/I)]$ for relative nontarget levels of -10, 0, and 10 dB, respectively, in order to maintain a particular value of the Weber fraction.

A. Methods

A two-interval, forced-choice procedure was used in which one interval of each trial contained the target pulse at its pedestal level and the other interval contained the target pulse at its pedestal level plus an increment. The listener's task was to select the interval containing the incremented target pulse. In conditions involving a nontarget pulse, it was presented in both intervals at a fixed level. In separate conditions, the level of the nontarget pulse was fixed at -10, 0, or 10 dB relative to the pedestal level of the target (i.e., 65, 75, or 85 dB SPL). JNDs were measured for the first pulse as target and, separately, for the second pulse as target. Intensity JNDs were also measured for a single pulse in isolation. The pedestal level of the target was 75 dB SPL in all conditions.

JNDs were measured adaptively using a three-down one-up tracking rule that estimated the 79.4% correct point on the psychometric function (Levitt, 1971). In each adaptive run, the step size of the increment was 4 dB, in units of

$10\log(\Delta I/I)$, until four reversals were obtained, after which the step size was changed to 2 dB until a total of 12 reversals were obtained. The average of the increments at which the last eight reversals occurred was calculated as the JND estimate for each adaptive run. Four such estimates were averaged to obtain a final estimate of the listener's intensity JND in each condition.

Both the target and nontarget pulses had the same parameters as those used in the previous experiment: 1-kHz tones, 10 ms in duration, with 5-ms on–off ramps (no steady state). Within each interval, the target and nontarget pulses were separated by the IPI, which ranged from 2–256 ms and which was defined as the time between the offset of the first pulse and the onset of the second pulse. The two intervals of each trial were separated by 500 ms of silence.

Four listeners from the previous experiment (all except S3) participated in this experiment. All conditions of the experiment were run in a pseudorandom order chosen by the experimenter.

B. Results and discussion

The difference between the intensity JNDs for each two-pulse condition and the one-pulse condition was computed for each listener in units of $10\log(\Delta I/I)$. The form of these differences was similar for all listeners and thus were averaged across listeners. Figure 6 shows these average differences, with the top and bottom panels representing data for conditions in which the first or second pulse, respectively, was the target. The symbol type indicates the level of the nontarget pulse with respect to the target pedestal level. The points plotted separately on the left-hand side of the panels indicate the threshold differences that would be predicted if perfect integration across the two pulses occurred.

Despite substantial variability in the actual size of JND differences between subjects, it can be seen that JNDs increased with increasing nontarget level roughly independent of IPI. In most cases, particularly for the −10 and +10 dB nontargets, the actual JNDs could not be predicted by assuming perfect temporal integration of the pulses. JNDs were higher than predicted for the −10 dB nontargets and they were lower than predicted for the +10 dB nontargets. Although JNDs depended upon the level of the fixed nontarget component, the fact that they did not change with IPI over the range studied here indicates that changes in weighting strategy across IPI that were seen in experiment 1 cannot be accounted for by differential sensitivity to intensity changes at different IPIs.

Plack *et al.* (1995) presented results for a similar task (their experiment 1), in which intensity JNDs were measured for a 30-ms tone in the presence of a second 30-ms masker tone that either preceded or followed the signal for IPIs ranging from 12.5 to 200 ms. They also found that thresholds were higher in all cases for the two-pulse condition relative to a one-pulse condition, but thresholds decreased slightly with decreasing IPI. No such decrease was found in the present results. In the Plack *et al.* experiment, the level of the nontarget pulse was 80 dB SPL and the level of the target pedestal was 50 dB SPL. It is possible that as the IPI was reduced in their experiment, the pedestal and increment were sufficiently masked by the nontarget pulse that the intensity discrimination task became more like a detection task. Consistent with this explanation, in subsequent experiments, Plack *et al.* found that intensity discrimination thresholds increased as the level of a second pulse that was 12.5 ms from the target was reduced. It should be noted, however, that the stimuli used by Plack *et al.* were more complicated than those in the present experiment, with a third pulse more temporally remote from the target.

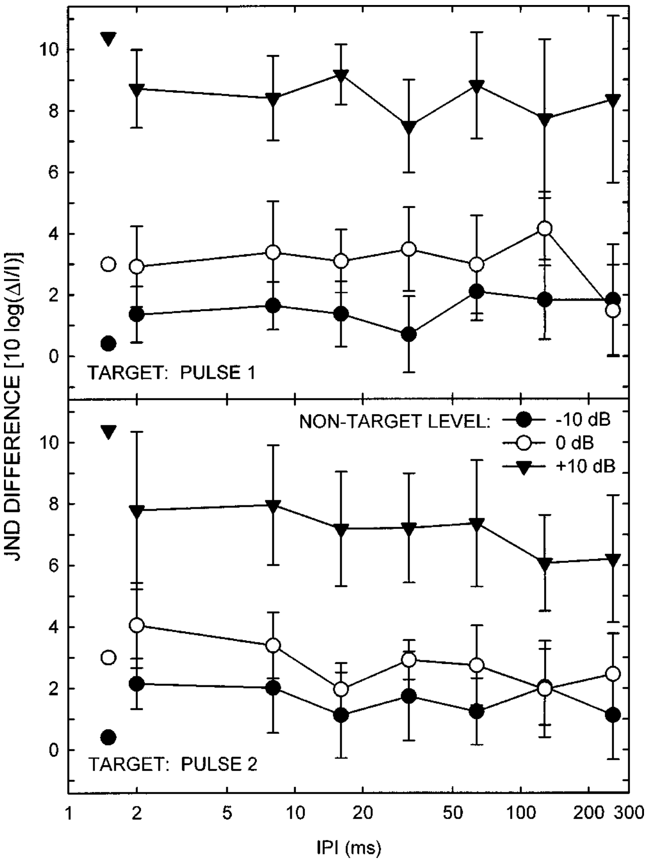

FIG. 6. The differences between intensity discrimination thresholds in the two-pulse conditions and the threshold for one pulse, plotted as a function of IPI. Different symbol types indicate the level of the nontarget pulse with respect to the level of the target pedestal, which was always 75 dB SPL. The upper panel shows data for conditions in which the first pulse was the target, and the lower panel shows data for the second pulse as the target. Each separate symbol at the left of the panels indicates the threshold increase that would be predicted for each nontarget level assuming perfect temporal integration of the pulses. The data are averaged across four listeners. Error bars represent standard errors of the mean.

Plack *et al.* accounted for their results in terms of a referential coding hypothesis, whereby the listener can make use of the second fixed pulse as an intensity reference and can encode the intensity of the target pulse in terms of its intensity relative to the reference rather than in terms of its absolute intensity. Indeed, in the present experiment, when the nontarget pulse was equal in level to the target pedestal, on each trial the listener was presented with two pairs of pulses, or three pulses of equal level and one pulse with an incremented level. It would seem that in the equal level (0-dB) condition, the listener could have used the fixed pulses as additional references, essentially getting additional ‘‘samples’’ of the pedestal level. However, even at the longest IPI (256 ms), thresholds in the two-pulse condition were elevated relative to the one-pulse condition, which itself contains two pulses separated by 500 ms on each trial. This suggests that intensity discrimination thresholds in a one-pulse condition, in which one pulse is presented in each of two intervals, would increase as the time between the intervals is decreased from 500 to 256 ms.

A number of studies have addressed the topic of interference between two intervals in an intensity discrimination task as a function of the time between the intervals (Tanner, 1961; Sorkin, 1966; Taylor and Smith, 1975). In general, it has been found that performance in a two-interval, two-alternative forced-choice task is maximal for an interstimulus interval (ISI) of about 500 ms. At shorter ISIs, it is assumed that the close temporal proximity of the intervals creates interference in short-term memory, limiting processing of either interval. For ISIs longer than about 500 ms, the memory trace of the first interval presumably decays or becomes increasingly corrupted by noise (Kinchla and Smyzer, 1967) which limits the accuracy of comparisons between the intervals. None of these hypotheses nor the present results imply simple temporal integration across stimuli for temporal separations greater than a few milliseconds. Rather, it is suggested that interference occurs between the information provided in the discrete temporal intervals at a central level of processing.

III. GENERAL DISCUSSION

A. Multiple-looks hypothesis

Experiment 1 measured the weighting functions that describe the manner in which listeners use short-term looks at the level of a stimulus in making decisions about changes in level at different temporal positions. In general, different weighting functions were computed when listeners were instructed to make different decisions about the stimuli even though the stimuli were statistically identical across trials. This indicates that listeners do in fact have access to level information in brief segments of a stimulus, and that listeners can make optimal use of that information to different degrees depending upon the temporal separation of the stimulus segments. The results of the difference condition and those of Lutfi (1990) and Buus (1999) suggest that it is easier for listeners to adopt an optimal weighting strategy when the task requires listeners to utilize the information in all of a group of short-term looks that are closely spaced in time. Listeners appear to have more difficulty selectively ignoring some of those looks, as in the target/variable conditions of experiment 1 and in Lutfi (1992). This may be true only for situations in which the looks are qualitatively similar, given the results of Viemeister and Wakefield (1991), which showed that changing the level of a noise burst temporally positioned between two tone bursts has no effect on detection thresholds for the two tone bursts.

As implied above, given that different weighting strategies are observed for different tasks, and that there is some variability in weighting strategies across listeners, it is possible that weighting strategies also would be affected by the type of signal. For example, the present experiments and those of Buus (1999) and Viemeister and Wakefield (1991) utilized signals consisting of temporally discrete tonal pulses. Although listeners in most cases used a nearly optimal weighting strategy when all of the discrete pulses were to be used in performing a listening task (e.g., detection of the pulses or comparison of the pulse levels), nonoptimal strategies were observed when listeners were instructed to base decisions on one pulse while ignoring the other (as in the target/variable conditions). Although detection of a signal consisting of a series of tone pulses yielded optimal weighting strategies, it is possible that a nonoptimal weighting strategy would be used when the task requires the listener to detect a stimulus consisting of a single long-duration tone. This was observed by Gilkey and Robinson (1986) and posited by Viemeister and Wakefield (1991) to account for previous temporal integration data. In addition, the weighting strategy is influenced by the relative statistics of the parameters of the stimulus elements (Berg, 1990; Lutfi, 1992) and by the characteristics of off-frequency maskers (Buus, 1999). It is clear that no single weighting strategy can account for performance with all stimuli and listening tasks, but that the weighting of short-duration segments of a stimulus can account for a substantial portion of responses in many specific tasks.

B. Comparison to binaural data

In an experiment that was a binaural analog to experiment 1 of the present study, Stellmack *et al.* (1999) measured weighting functions for two temporally separated clicks with different interaural delays. In that study as in the present paper, listeners were instructed to respond to the information in only one stimulus element or to compare the stimulus elements to one another. Although the experiments were performed with mostly different listeners and slightly different stimuli, a comparison of the data from the two experiments suggests some possible similarities between the weighting functions measured in the two tasks. For example, in the binaural task in which the interaural delays of both clicks varied from trial to trial, percent correct on average was lowest for an intermediate range of interclick intervals (8–32 ms). In addition, as in the present study, there was evidence that performance was limited by two different factors at short and long interclick intervals: the precedence effect at short interclick intervals (see Litovsky *et al.*, 1999, for a recent review), and an inability to discern the temporal order of the interaural delays at longer interclick intervals, a limitation found in a number of sensory modalities by Hirsh and Sherrick (1961). Additional data would need to be gathered in both the monaural and binaural conditions from the same listeners in order to establish whether weighting functions at the larger range of IPIs might be driven by a common limitation on the ability to judge temporal order.

C. Conclusions

The weighting functions obtained in experiment 1 suggest that the auditory system treats long-duration stimuli as if they are composed of many short-term looks, with the information in each look weighted in a particular way depending upon the task at hand. It appears that when information-bearing, qualitatively similar stimulus elements are separated by less than about 16 ms, listeners adopt a weighting strategy that is closer to optimal when required to make decisions based on a combination of all of the looks rather than on a subset of the looks. Listeners display an increasing ability to ignore particular temporal segments of the stimulus with increasing temporal spacing above about 16 ms.

Intensity JNDs for one pulse are elevated when a second fixed-level pulse is present, but cannot be predicted by assuming perfect temporal integration across the pulses. The amount of JND elevation is dependent upon the level of the fixed-level pulse but not on the temporal spacing of the pulses over the range from 2 to 256 ms. This indicates that the various weighting strategies measured in experiment 1, which differed across temporal separations, are not attributable to decreased sensitivity to level changes in one pulse when a second pulse is present.

ACKNOWLEDGMENTS

The authors would like to thank the associate editor (Dr. Sid P. Bacon) and two anonymous reviewers for their careful reviews and helpful suggestions. The authors also thank Dr. Martin E. Rickert and Dr. Magda Wojtczak for fruitful discussions of the material herein. This work was supported by Research Grant No. R01 DC 00683 from the National Institute on Deafness and Communication Disorders, National Institutes of Health.

[1] Using the equations of Richards and Zhu (1994), one can estimate the relative magnitudes (in proportion to the relative weights) of the variance of the continuous decision variable (σ_D^2) and the variance of the additive internal noise process (σ_ε^2). From these values, one can estimate the proportion of variance in the listener's responses attributable to the additive noise process ($\sigma_\varepsilon^2/\sigma_D^2$), and the proportion of variance attributable to the sum of the weighted level changes ($1-\sigma_\varepsilon^2/\sigma_D^2$). A graph describing the latter quantity as a function of IPI is extremely close in form to that describing the proportion of responses predicted by the weights as a function of IPI, which is used in the present paper. Because relative comparisons are made across conditions in the present manuscript, either measure would provide the same information. The proportion of responses predicted by the weights was also chosen for its intuitive appeal.

[2] The criterion, k, in Eq. (2) was estimated as described by Richards and Zhu (1994). In general, the estimated criteria were very close to zero (the unbiased value of the criterion). When the estimated criterion was used to predict the listener's responses, very few additional responses were predicted compared to when an unbiased criterion ($k=0$) was assumed. In the most extreme case, 33 additional responses out of 1000 were correctly predicted when the estimated criterion was used. In most cases, fewer than ten additional responses were predicted using the estimated criterion, indicating that listener's responses for all practical purposes were unbiased.

Babkoff, H., and Sutton, S. (**1963**). ''Perception of temporal order and loudness judgments for dichotic clicks,'' J. Acoust. Soc. Am. **35**, 574–577.

Berg, B. G. (**1989**). ''Analysis of weights in multiple observation tasks,'' J. Acoust. Soc. Am. **86**, 1743–1746.

Berg, B. G. (**1990**). ''Observer efficiency and weights in a multiple observation task,'' J. Acoust. Soc. Am. **88**, 149–158.

Buus, S. (**1999**). ''Temporal integration and multiple looks, revisited: Weights as a function of time,'' J. Acoust. Soc. Am. **105**, 2466–2475.

Buus, S., Zhang, L., and Florentine, M. (**1996**). ''Stimulus-driven, time-varying weights for comodulation masking release,'' J. Acoust. Soc. Am. **99**, 2288–2297.

de Boer, E. (**1985**). ''Auditory time constants: A paradox?,'' in *Time Resolution in Auditory Systems*, edited by A. Michelsen (Springer-Verlag, Berlin), pp. 141–158.

Forrest, T. G., and Green, D. M. (**1987**). ''Detection of partially filled gaps in noise and the temporal modulation transfer function,'' J. Acoust. Soc. Am. **82**, 1933–1943.

Gilkey, R. H., and Robinson, D. E. (**1986**). ''Models of auditory masking: A molecular psychophysical approach,'' J. Acoust. Soc. Am. **79**, 1499–1510.

Green, D. M. (**1960**). ''Auditory detection of a noise signal,'' J. Acoust. Soc. Am. **32**, 121–131.

Green, D. M. (**1973**). ''Temporal acuity as a function of frequency,'' J. Acoust. Soc. Am. **54**, 373–379.

Green, D. M. (**1985**). ''Temporal factors in psychoacoustics,'' in *Time Resolution in Auditory Systems*, edited by A. Michelsen (Springer-Verlag, Berlin), pp. 122–140.

Green, D. M., and Swets, J. A. (**1966**). *Signal Detection Theory and Psychophysics* (Wiley, New York).

Hirsh, I. J. (**1959**). ''Auditory perception of temporal order,'' J. Acoust. Soc. Am. **31**, 759–767.

Hirsh, I. J., and Sherrick, Jr., C. E. (**1961**). ''Perceived order in different sense modalities,'' J. Exp. Psychol. **62**, 423–432.

Kinchla, R. A., and Smyzer, F. (**1967**). ''A diffusion model of perceptual memory,'' Percept. Psychophys. **2**, 219–229.

Leek, M. R., and Watson, C. S. (**1984**). ''Learning to detect auditory pattern components,'' J. Acoust. Soc. Am. **76**, 1037–1044.

Levitt, H. (**1971**). ''Transformed up-down methods in psychophysics,'' J. Acoust. Soc. Am. **49**, 467–477.

Litovsky, R. Y., Colburn, H. S., Yost, W. A., and Guzman, S. J. (**1999**). ''The precedence effect,'' J. Acoust. Soc. Am. **106**, 1633–1654.

Lutfi, R. A. (**1990**). ''Informational processing of complex sound. II. Cross-dimensional analysis,'' J. Acoust. Soc. Am. **87**, 2141–2148.

Lutfi, R. A. (**1992**). ''Informational processing of complex sound. III: Interference,'' J. Acoust. Soc. Am. **91**, 3391–3401.

Lutfi, R. A. (**1995**). ''Correlation coefficients and correlation ratios as estimates of observer weights in multiple-observation tasks,'' J. Acoust. Soc. Am. **97**, 1333–1334.

Plack, C. J., Carlyon, R. P., and Viemeister, N. F. (**1995**). ''Intensity discrimination under forward and backward masking: Role of referential coding,'' J. Acoust. Soc. Am. **97**, 1141–1149.

Richards, V. M., and Zhu, S. (**1994**). ''Relative estimates of combination weights, decision criteria, and internal noise based on correlation coefficients,'' J. Acoust. Soc. Am. **95**, 423–434.

Ronken, D. A. (**1970**). ''Monaural detection of a phase difference between clicks,'' J. Acoust. Soc. Am. **47**, 1091–1099.

Schlauch, R. S., DiGiovanni, J. J., and Ries, D. R. (**1999**). ''Cognitive limitations on auditory temporal resolution,'' Association for Research in Otolaryngology Mid-Winter Meeting, February 1999.

Sorkin, R. D. (**1966**). ''Temporal interference effects in auditory amplitude discrimination,'' Percept. Psychophys. **1**, 55–58.

Stellmack, M. A., Dye, Jr., R. H., and Guzman, S. J. (**1999**). ''Observer weighting of interaural delays in source and echo clicks,'' J. Acoust. Soc. Am. **105**, 377–387.

Tanner, Jr., W. P. (**1961**). ''Physiological implications of psychophysical data,'' Ann. (N.Y.) Acad. Sci. **89**, 752–765.

Taylor, M. M., and Smith, S. M. (**1975**). ''Monaural detection with contralateral cue. V. Interstimulus interval in MDCC and amplitude discrimination,'' J. Acoust. Soc. Am. **57**, 1500–1511.

Viemeister, N. F. (**1979**). ''Temporal modulation transfer functions based upon modulation thresholds,'' J. Acoust. Soc. Am. **66**, 1364–1380.

Viemeister, N. F., and Wakefield, G. H. (**1991**). ''Temporal integration and multiple looks,'' J. Acoust. Soc. Am. **90**, 858–865.

Wright, B. A., and Dai, H. (**1999**). ''Detection of sinusoidal amplitude modulation at unexpected rates,'' J. Acoust. Soc. Am. **104**, 2991–2996.